Power Lines, Line Transects, and GIS

J Wanzer Drane, PE, PhD (1),* Heidi L Weiss, PhD (2), Tim E Aldrich, MPH, PhD (3), Dana L Creanga, PhD (4), Gerald F Pyle, PhD (5)

(1) School of Public Health, University of South Carolina, Columbia, SC; (2) Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL; (3) Department of Health and Environmental Control, Centers for Disease Control and Prevention, US Department of Health and Human Services, Columbia, SC; (4) Computer Horizons, Indianapolis, IN; (5) Professor of Geography, University of North Carolina at Charlotte, Charlotte, NC

Abstract

With the advent of address matching, line and band transects can be used to take useful random samples of homes, workplaces, or other such objects in an area. In this study of power lines and cancers in a part of North Carolina, we created band transects by adding buffers to each side of a line transect. The band transects were used to sample homes of cancer victims in order to estimate the density of new cases within a fixed geographic region. The entire set of new cases over one year defines the sampling frame. Trunk power lines were the assumed hazard. Therefore, the band transects were modified by removing a center strip representing the rights of way of the power lines. The intersections of addresses of cancer victims with buffer zones outside the rights of way created a null distribution of cancer densities within the study area. The power line rights of way, with buffers added to each side, were then intersected with the frame of addresses to provide a sample of affected persons. This paper presents four major conclusions of the study. First, the distributions of the numbers of cancer cases were very well modeled by Poisson distributions. Second, 40-meter-wide buffers were more efficient at capturing cases than were those 20 and 70 meters wide. Third, the density of cancer cases within the buffers of power lines was approximately half that within the randomly created band transects. Finally, a case is made for the continued development of this methodology. From this experience was born the concept of "the shadow of the hazard," an appropriately shaped quadrat that represents a hazard's area of influence and can be used to sample the affected and unaffected portions of a population.

Keywords: linear hazards, cancer, spatial statistics, sampling, prevalence

The Line Transect Method

Background

The line transect and its extension, the band transect, are normally used to estimate D, the population density of wildlife in an area. (D is the number of subjects per unit area.) The technique's application consists of walking a straight path of random or fixed length, L, through the region of interest. The observer records the number of subjects identified and their estimated right angle distances, y, from the line transect when first observed. This method of sampling is well documented (1–8). Let the function

* J Wanzer Drane, Professor of Biostatistics, University of South Carolina, School of Public Health, Green and Sumter Streets, Columbia, SC 29201 USA; (p) 803-777-5053; (f) 803-777-2524; E-mail: Wdrane@sph.sc.edu

$g(y)=P(\text{observing an object} \mid y)$ (1). It is assumed that $g(0)=1$, and $g(\infty)=0$. That is, a subject exactly on the line is seen with probability equal to 1, and at great distances it will not be seen at all. During the traverse, n subjects will be seen at distances Y_i, $i=1, \ldots, n$. Let $f(y)$ denote the probability density function of Y.

Estimation of the Density of Cancer Cases Using the Line Transect Method

Let the object being observed be the residence of a cancer victim. Let $g(y)$ and $f(y)$ be defined as above. D, instead of the number of wildlife subjects, is the number of cancer cases per unit area. Consider a random line of length L with a fixed observation width W within a study area. The band has width $2W$, length L, area $2LW$, and a number n of cases within it. The estimate of D is given by $\hat{D} = \frac{n}{2L\mu}$, where $\mu = \int_0^W g(y)dy$ In this application, $g(y)=1$ for all y, and $0 \le y \le W$, because every cancer case within a band will be detected and counted with certainty. Using a geographic information system (GIS), precise coordinates of the home of each cancer case are encoded. Thus, the estimate of cancer case density is $\hat{D} = \frac{n}{2LW}$, which is simply the count divided by the area of the band transect.

Statistical Properties of the Estimate of D

In typical transect studies, replicate random lines are created over the entire study area. Let there be R replicate random lines with lengths h_i, with respective counts n_i, where $i=1, 2, \ldots, R$. The estimate of D can be calculated as above for each individual transect or as $\hat{D} = \frac{\sum n_i}{2W\sum h_i}$, where the summations are over the R replicates. This estimate is the maximum likelihood estimate of $\hat{D}$ (3). We will estimate the variance of D directly using replicate transects and also using the ever-popular jackknife.

Spatial Analysis of Disease and Environmental Hazards

Here we are interested in describing and analyzing disease patterns in physical space, defined, for example, by longitude and latitude or their transformations of location and scale. As an example, many epidemiological studies have focused on residential exposures to electromagnetic fields created by the passage of electrical currents through power lines (9–16). Residential proximity to environmental factors, presumed to be hazards, has been central to some of those studies. In the same studies, moreover, maps showing power lines near homes—or, for other hazards, similarly indicative maps showing residential proximities to the presumed hazards—were sufficient to indicate exposure to environmental hazards.

It is therefore beneficial for the inquiry into the effects of distance between potential environmental hazards and cancer occurrence to capitalize on the advances in GIS technology. This study demonstrates the application of the line transect method, by way of GIS, to this inquiry, the purpose of which is to estimate the densities of cancer cases within proximities of linear objects. The linear objects in our study were the electrical trunk lines serving a large city, Charlotte, North Carolina.

Background on Geographic Information Systems

A GIS is any manual or computer-based set of procedures used to store and manipulate geographically referenced data (17). A GIS should be capable of data input, data management, and manipulation and analysis of geographically referenced data. A distinction should be made between cartographic systems and GIS. The main function of a cartographic system is to store maps in automated form and generate computer-based maps. A GIS should have additional capabilities—it should be able to integrate layers of geographically referenced data, perform analysis on these layers of data, and predict or evaluate spatial relationships and outcome phenomena. A GIS layer is a set of data with geographic references compatible in scale and location with cartographic features stored and manipulated by the GIS. Layers include, but are not limited to, spatial attributes such as longitude and latitude, area, length, use, population within polygons (e.g., city blocks), or boundary designations. Layers can also include population densities, socioeconomic status variables, census information, or personal information together with the coordinates of the place where a person lives or works. The ability to explore the interactions or interrelationships between geographical variables and geographically referenced other variables makes GIS potentially a very powerful tool for geographic analysis.

The area of geographic health information systems is an emerging area of GIS applications. GIS has been used for the planning and delivery of health services (18). Gisler (19) cited the use of GIS for assessing environmental risks as one of the future research directions of spatial analysis of diseases. GIS facilitates the application of spatial techniques in generating hypotheses about environmental hazards and cancer risks. Furthermore, GIS provides an initial impression of the relationship between cancer cases and potential environmental hazards—a starting point from which to analyze data from typical state cancer registries and surveillance programs.

Materials and Methods

Sample

The line transect method of sampling and ARC/INFO (20), a GIS, were applied to data from Mecklenburg County, North Carolina. This study was limited to the area covered by the Dual Independent Map Encoding (DIME) files. These datasets, called coverages, included digitized spatially related data on homes of cancer victims, roads, streams, power lines, census tracts, and street addresses. Each coverage has a corresponding attribute table that describes the geographic features within the area. The US Census Bureau's TIGER/Line files (21) were used to create the coverages. TIGER files provide digital data on all census geographic boundaries, codes, longitude and latitude coordinates, feature names, addresses, and zip codes. Mecklenburg County was divided into sub-areas, and only the sub-areas defined by the DIME files were investigated. This ensured address matching of all cancer cases. The North Carolina Central Cancer

Registry's population-based cancer incidence database for 1990 supplied the incidence data on cancer for this study (22,23).

ARC/INFO was used to perform a cross-sectional geographic analysis by integrating these coverages. Addresses of newly diagnosed cancer cases for 1990 were represented as points, roads and power lines as arcs, and census tracts as polygons. To perform spatial analysis, ARC/INFO was used to create random transects across the county, create buffers of specified widths around the transects, overlay the resulting buffered transects with other coverages such as cancer case coverage or census tract coverage, and calculate such measures as the number of cancer cases within the transect and the total area of the band transects created from the line transects and their respective buffers.

Spatial Sampling Procedures

The geographic area considered in this study is the DIME area of Mecklenburg County, North Carolina. The population under investigation is the newly diagnosed cancer cases for 1990 who live within that DIME area. The cancer cases are represented as points whose longitude and latitude coordinates locate them within the area. Those cancer case points constitute the cancer case coverage or layer, along with an attribute table associated with them. That attribute table contained ID, RACE, AGE, GENDER, ZIPCODE, ADDRESS, CENSUS TRACT, and ICDOCODE (type of cancer). An address coverage was used to create the cancer case coverage. Within it were stored addresses and corresponding IDs. Address matching was used to obtain the coordinates for each cancer case address. Another coverage was composed of the census tracts and their associated polygon attribute table. This attribute table included data on census tract number, census tract ID, and population size.

Line transects with their respective buffers were the sampling units. A simple random sample of lines throughout the entire county was drawn. This was done by creating a random number of random point pairs $[(x_1, y_1), (x_2, y_2)]$ using a random number generator. The resulting data were read into ARC/INFO. Using the program's GENERATE command, line segments were drawn between the pairs of points. Those segments were considered to be one replicate. This process was repeated 10 times, so 10 replicates were drawn. Each replicate consisted of, at most, 20 line transects.

The lengths of the transects were calculated and a 30-meter corridor was drawn around each. These corridors represented the rights of way beneath the power lines, where there could be no houses. Beyond the corridors, buffers were created of widths 20, 40, and 70 meters. These buffers of varying width acted as surrogates for exposure to electromagnetic fields. The resulting coverage of transects and buffers was overlaid on the cancer case coverage. The cancer cases denoted as points within the buffered transects were elements within the sampling units. Note that the lengths of the transects differed, and many transects intersected one another.

Transects from a single replicate were pooled to provide an estimate of the density of cancer cases. Ten estimates of the density of cancer cases were calculated, one for each of the ten replicates. From these an empirical distribution of the density could be calculated and then used to calculate an overall estimate, $\boldsymbol{D_0}$. The next step was to create line transects using a geographic feature in the environment. The high-power electrical trunk lines were the linear feature of choice. The entire segment of the trunk lines feeding the greater Charlotte area within the Mecklenburg County DIME area was considered a single transect. Using the same procedure that was used for the random line transects, a single estimate of the density of cancer cases was calculated. A test of equality between the density of cancer cases along the random transects and the density along the power line transect was carried out. In both cases, the number of cases found by the buffers followed a Poisson probability mass function with a mean of $2LWD$, and under H_0, $\boldsymbol{D_0}$ is substituted for D (1). Denoting $\boldsymbol{D_p}$ as the density near the power lines, n_p was the Poisson test statistic with mean (parameter) $\Theta_0 = 2WLp\boldsymbol{D_0}$. That is,

$$f(n) = P(n \mid \Theta_0) = \frac{e^{-\Theta_0}\Theta_0^n}{n!},\ \mathrm{n} = 0,1,2,\ \ldots,\ \infty .$$

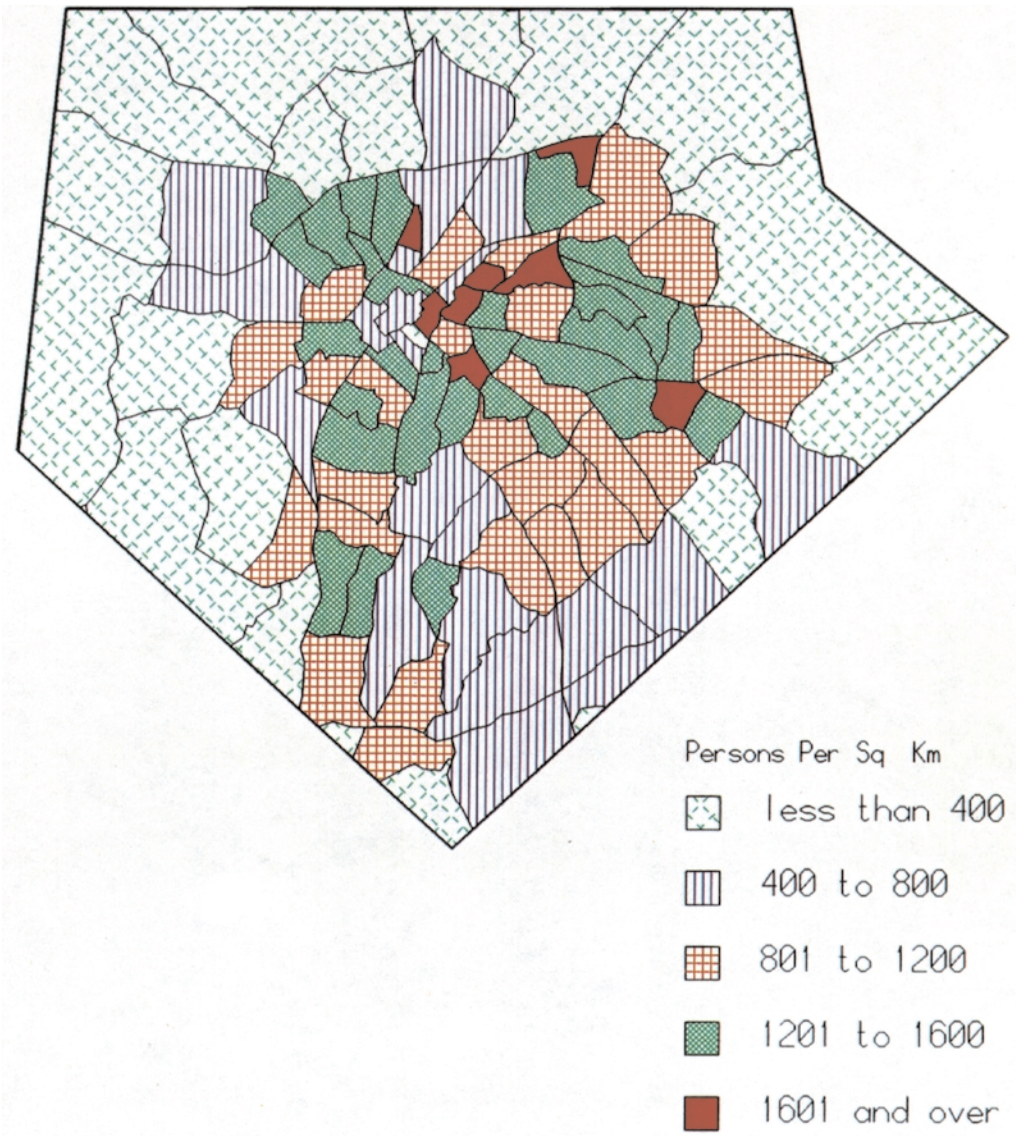

Figure 1 Population density by census tract, DIME file area of Mecklenburg County, North Carolina.

The p-values were calculated using the minimum likelihood method (24). The above procedure was done for *W*=20, 40, and 70 meters.

Results

Estimation of Cancer Density Using Random Line Transects

The final coverages used in the analysis include the census tracts (Figure 1), the high-power transmission lines (Figure 2), and the roads (Figure 3), all within the DIME area. The result of address matching is shown in Figure 4.

Each replicate, consisting of many line transects, formed a new coverage. For each replicate, a new analysis was done as follows. The coverage was clipped to fit within the boundary of the DIME area. Right-of-way corridors of 30 meters were drawn around the line transects (Figure 5), and buffers of 20, 40, or 70 meters were drawn outside the corridors (Figure 6). The corridors were then erased and the new buffered coverage was overlaid on the cancer case coverage (Figure 7).

The areas of the buffers and the numbers of cancer cases within them were

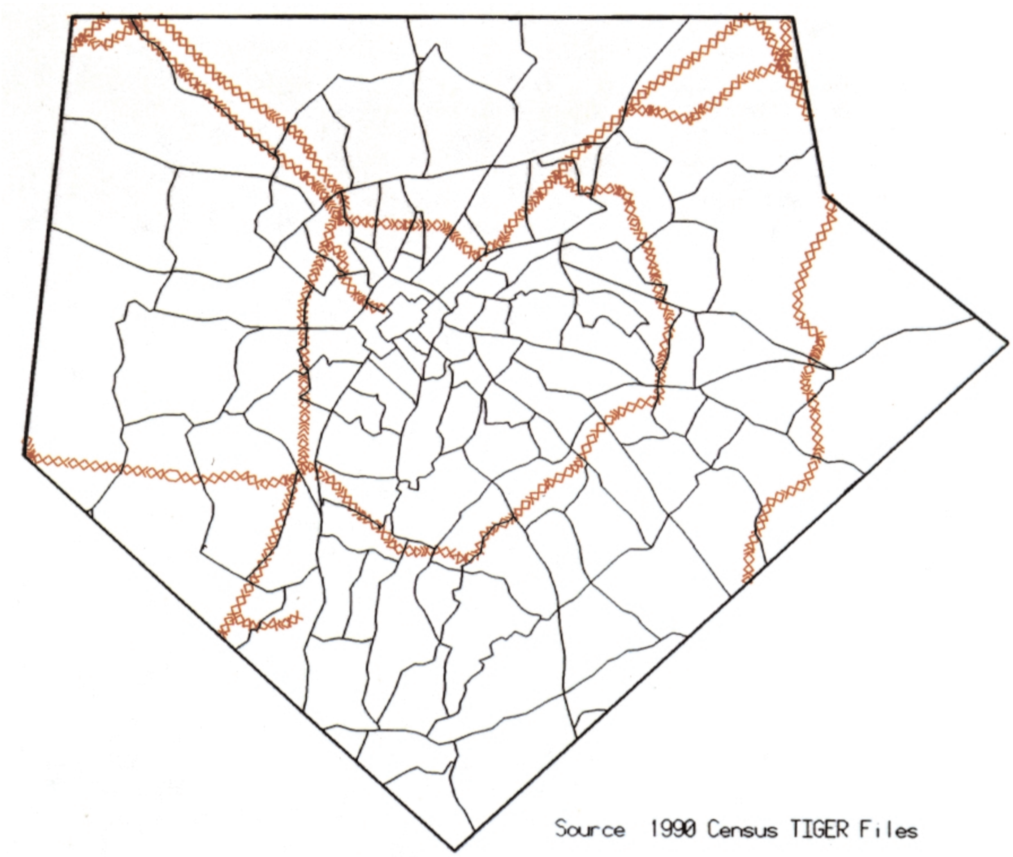

Figure 2 High-power transmission lines, DIME area of Mecklenburg County, North Carolina. Source: (21)

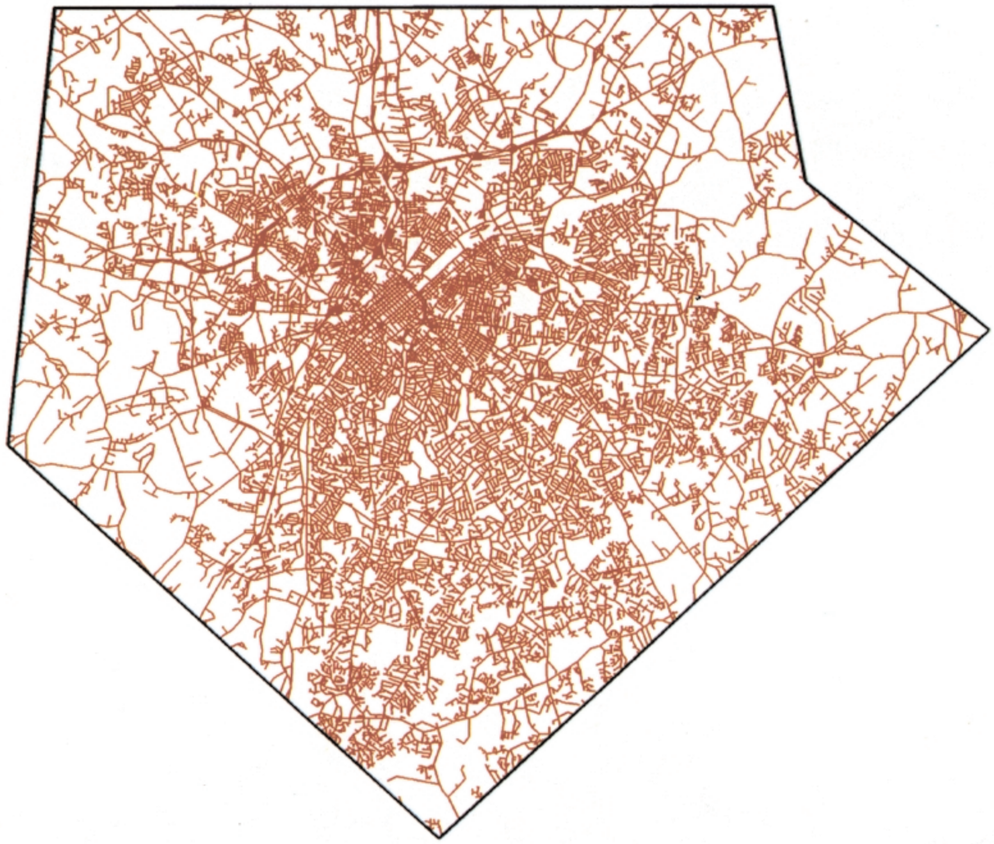

Figure 3 Roads coverage, DIME area of Mecklenburg County, North Carolina. Source: (21)

calculated using ARC/INFO. Note that the buffers are not entirely rectangular, and their areas are therefore not exactly 2*LW*. ARC/INFO calculates the areas accurately, taking into account the rounded or otherwise altered ends of the buffers.

The first analysis was done for a buffer 20 meters wide, and the results of the density estimates for that and the two other buffer widths were stored in a table (Table 1). These 10 replicates provide an empirical distribution of cancer cases within the buffered transects. The combined estimates of D_0, shown in Table 2, are 1.72, 1.77, and 1.63 cases per square kilometer, respectively, for buffers 20, 40, and 70 meters wide. Table 2 also shows estimates of the variances of the estimates of cancer densities.

Estimates of Cancer Density Using Power Transmission Lines

The null hypothesis of $D_0 = D_p$ versus not equal was tested for each of the buffer widths; the corresponding p-values are given in Table 3. The several D_p shown in Table 4 were compared with their respective D_0 counterparts of Table 2. The p-values in each case gave no evidence to support a conclusion that D_p and D_0 were different.

At least for this study, which is *not* an epidemiological investigation, there is no

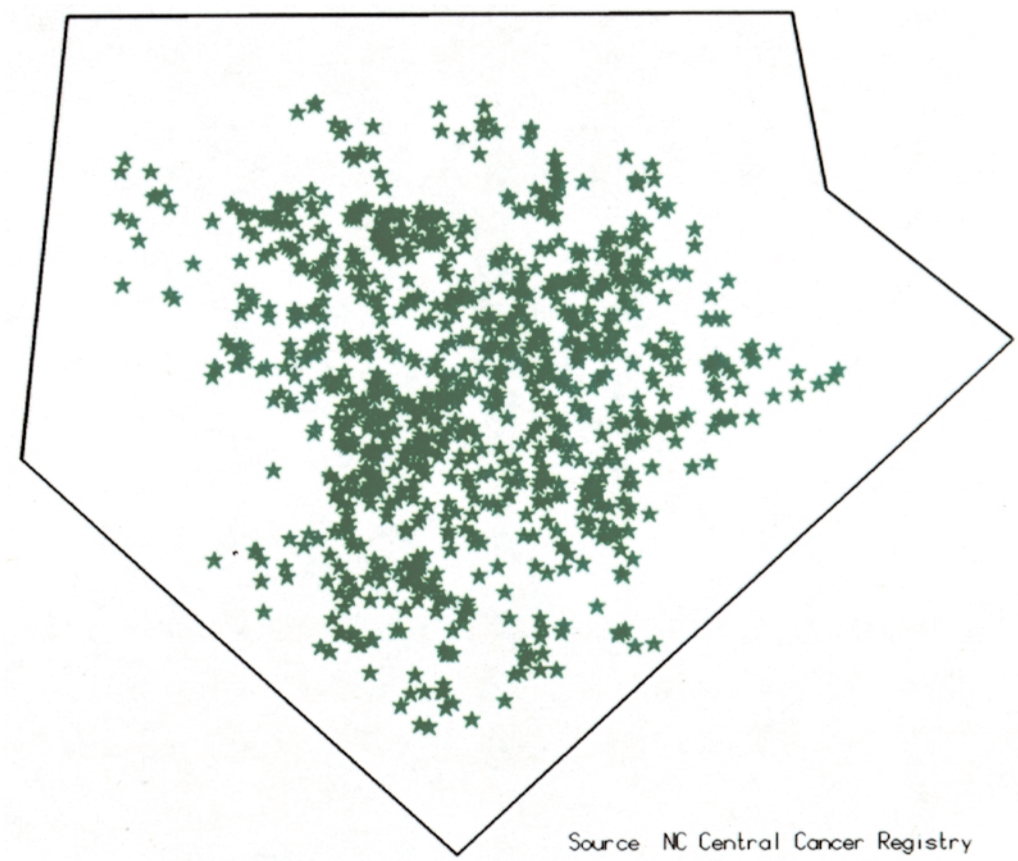

Figure 4 1990 cancer cases, DIME area of Mecklenburg County, North Carolina.
Source: (22,23)

evidence to implicate trunk power lines as a source of cancer formation. This will be addressed again in the discussion.

Discussion

The line transect method has not been previously applied to spatial analysis of health data, at least to the knowledge of these writers. This study's goal was to demonstrate the plausibility of the method's use in estimating cancer case density. This we have done. The densities of cancer cases near supposed hazards provide evidence to suggest further studies of situations in which the densities are higher than those in the non-exposed portions of the population. The methodology used in this study, however, does need refining.

GIS goes hand-in-glove with analyses of this type. Aside from making it possible to overlay various layers on a base map, GIS enables users to manipulate and analyze spatially indexed data. In this study, for example, GIS made it possible to generate random

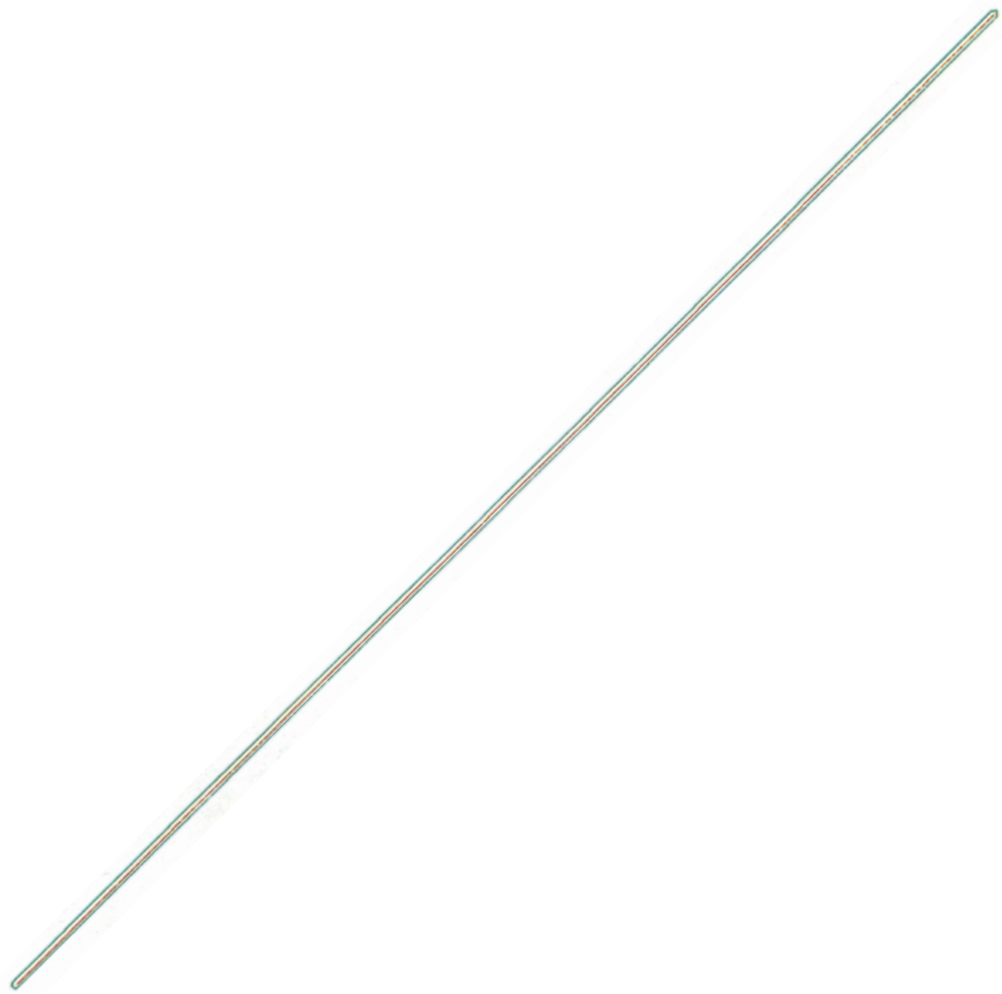

Figure 5 Random line transect with 30-meter corridor.

lines within a well-defined geographic area, remove corridors, add buffers outside the corridors, and overlay and intersect the point pattern of cancer cases with the transects, thereby providing estimates of cancer case densities.

Many insights can be garnered by applying GIS to such data. First, many existing surveillance programs could expand their reporting procedures to include such characteristics as occupations, industries, and time spent at home, at work, and in transit. The ability of GIS to integrate non-spatial data increases our ability to perform ecological analyses on human health data.

It should be pointed out again that this study was limited to hazards of a linear form such as power lines. The buffers outside linear hazards we call “shadows of the hazard.” This idea and name can be generalized to hazards of varying geometries and

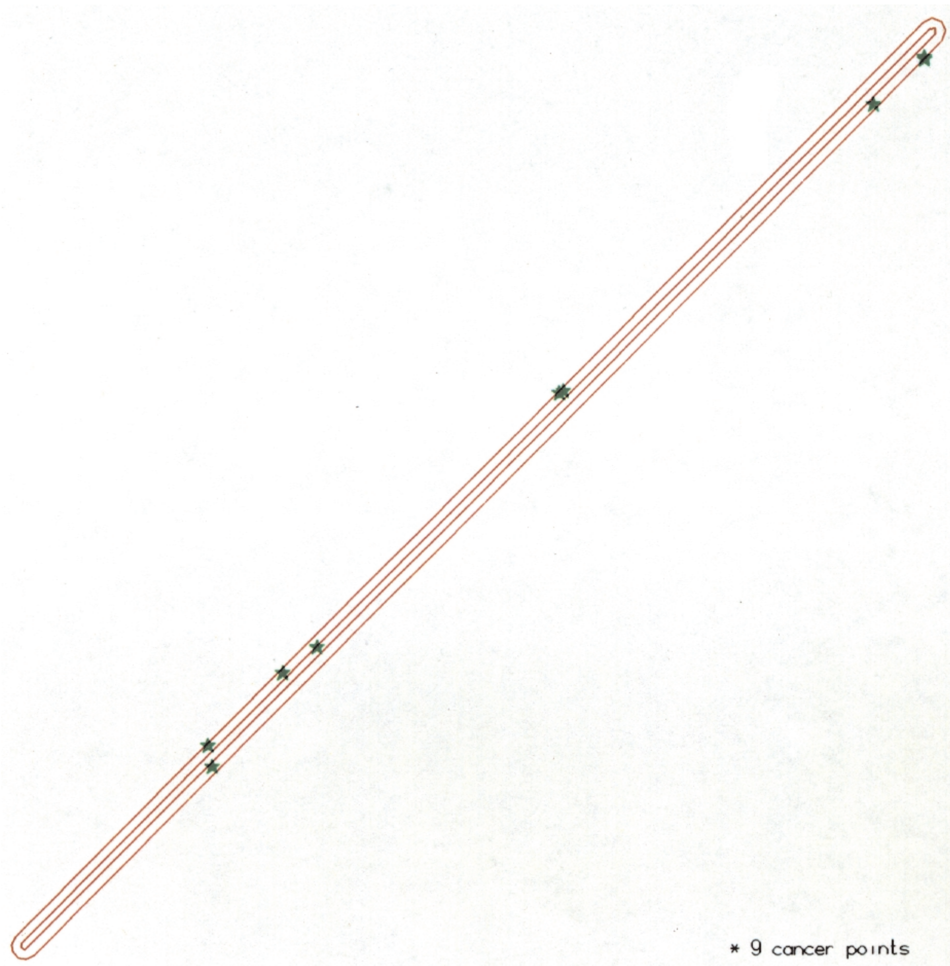

Figure 6 Random line transect with 70-meter buffer around 30-meter corridor.

sizes. Wind roses, for example, are shadows of the hazard of exhaust gases emitted to the atmosphere. A "cookie cutter" could be created from a wind rose with its core—which represents the hazard area itself—omitted. Randomly placed quadrats (the cookie cutter) would be used to sample areas and obtain densities, as we have done here with linear band transects with their cores (power line rights of way, in our case) removed.

Finally, it should be noted that we did not discriminate by cancer type, and population densities in the various census tracts were not taken into consideration. Now that our method has been demonstrated, you can improve on it at least in these two ways.

Acknowledgments

Thanks to Doctors C Macera, J Hussey, and S Weinrich for their contributions to this project; L Shirley, J White, and T White, who were resource persons for ARC/INFO; and T Weiss for his helpful editorial suggestions.

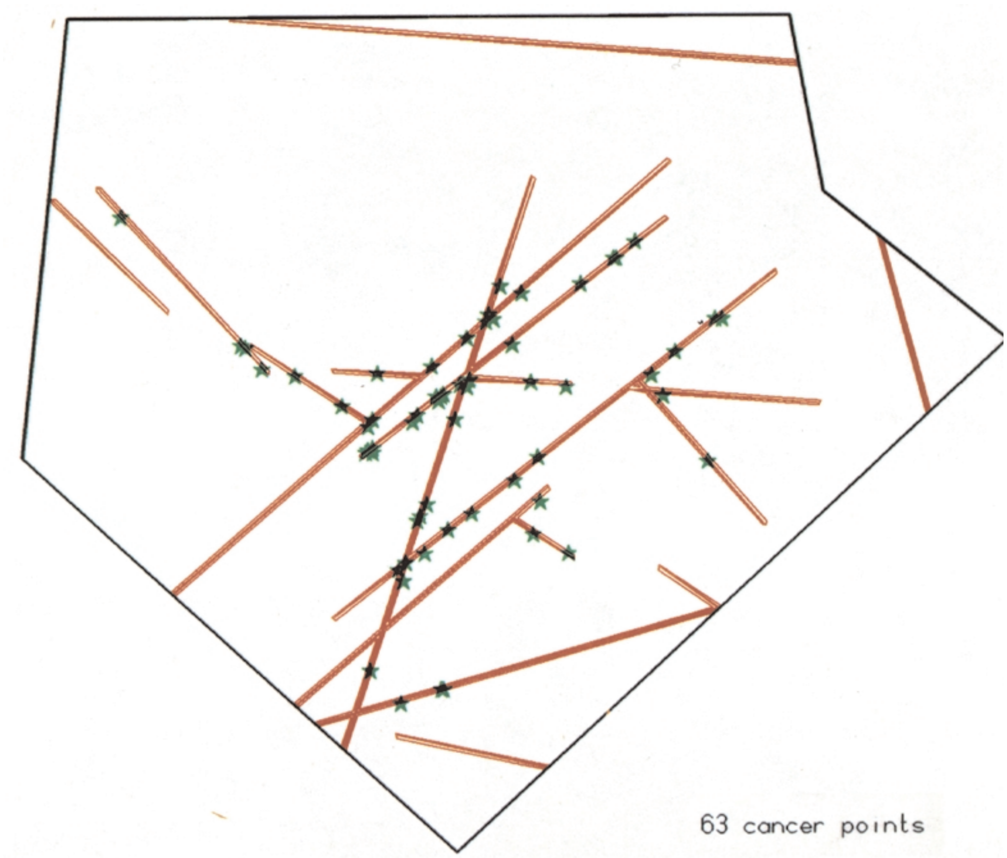

Figure 7 Random line transects and cancer cases within 70-meter buffers.

Table 1 Estimates of Cancer Case Density for Random Line Transects Using 20-, 40-, and 70-Meter Buffers

	Density Estimates per Square Kilometer		
Replicate	**20-Meter Buffer**	**40-Meter Buffer**	**70-Meter Buffer**
1	2.00	2.42	1.89
2	1.75	1.57	1.60
3	2.37	2.78	1.68
4	1.33	1.73	1.83
5	0.98	1.16	1.13
6	0.93	1.60	1.42
7	2.86	1.86	1.71
8	1.55	1.42	1.40
9	1.45	1.54	1.97
10	1.77	1.85	1.91

Table 2 Weighted Means of Estimates of D_0 and Variances of D_0

Buffer Width (meters)	D_0	Variance of D_0			
		Direct	Indirect	Jackknife	D_0/Area
20	1.72	0.043	0.037	0.039	0.023
40	1.77	0.028	0.019	0.015	0.012
70	1.63	0.008	0.007	0.007	0.006

Table 3 Tests of the Hypothesis that D_p=D_0

Buffer Width (meters)	D_0	D_p	p-value
20	1.72	0.53	0.427
40	1.77	0.53	0.432
70	1.63	1.15	0.999

Table 4 Estimates of Cancer Case Densities Near Power Lines Using 20-, 40-, and 70-Meter Buffers

Buffer Width (meters)	Number of Cases	Area (square kilometers)	D_p (cases per square kilometer)
20	3	5.64	0.53
40	6	11.34	0.53
70	22	19.99	1.15

References

1. Buckland ST. 1985. Perpendicular distance models for line transect sampling. *Biometrics* 41:177–95.
2. Upton GJG, Fingleton B. 1985. *Spatial data analysis by example; Point pattern and quantitative data, volume 1.* New York: John Wiley & Sons.
3. Burnham KP, Anderson DR, Laake JL. 1980. Estimation of density from line transect sampling of biological populations. *Wildlife Monograph No. 72, Supplement to Journal of Wildlife Management* 44.
4. Burnham KP, Anderson DR. 1976. Mathematical models for nonparametric inferences from line transect data. *Biometrics* 32:325–36.
5. Pollock KH. 1978. A family of density estimators for line transect sampling. *Biometrics* 34:475–8.
6. Eberhardt LL. 1968. A preliminary appraisal of line transects. *Journal of Wildlife Management* 32:82–8.
7. Gates CE, Marshall WH, Olson DP. 1968. Line transect method for estimating grouse population densities. *Biometrics* 24:135–45.
8. Seber GAF. 1973. *The estimation of animal abundance and related parameters.* London: Charles Griffin.

9. Wertheimer N, Leeper E. 1979. Electrical wiring configurations and childhood cancer. *American Journal of Epidemiology* 109:273–84.
10. Wertheimer N, Leeper E. 1982. Adult cancer related to electrical wires near the home. *International Journal of Epidemiology* 109:345–55.
11. Fulton JP, Cobb S, Preble L, et al. 1980. Electrical wiring configuration and childhood leukemia in Rhode Island. *American Journal of Epidemiology* 111:292–6.
12. McDowall ME. 1986. Mortality of persons resident in vicinity of electricity transmission facilities. *British Journal of Cancer* 53:271–9.
13. Tomenius L. 1986. 50-Hz electromagnetic environment and the incidence of childhood tumors in Stockholm County. *Bioelectromagnetics* 7:191–207.
14. Coleman MP, Bell CMJ, Taylor HL, Primic-Zakelj M. 1989. Leukaemia and residence near electricity transmission equipment: A case-control study. *British Journal of Cancer* 60:793–8.
15. Savitz DA, Wachtel H, Barnes FA, John EM, Tvrdik JG. 1988. Case-control study of childhood cancer and exposure to 60-Hz magnetic fields. *American Journal of Epidemiology* 123:21–38.
16. Stevens RK, Stevens RG, Kaune, et al. 1988. Acute nonlymphocytic leukemia and residential exposure to power frequency magnetic fields. *American Journal of Epidemiology* 123:10–20.
17. Starr J, Estes J. 1990. *Geographic information systems: An introduction.* Princeton: Prentice Hall.
18. Twigg L. 1986. Health-based GIS: Their potential examined in the light of existing data sources. *Social Science in Medicine* 30:963–73.
19. Gisler W. 1986. The uses of spatial analysis in medical geography: A review. *Social Science in Medicine* 23:963–73.
20. Environmental Systems Research Institute, Inc. (ESRI). 1991. *ARC/INFO version 5.* Redlands, CA: ESRI.
21. US Census Bureau. 1990. TIGER: *The coast-to-coast digital map data base.* Washington, DC: US Department of Commerce, Bureau of the Census. November.
22. North Carolina Central Registry. 1992. Cancer incidence in North Carolina: 1990 county specific numbers. Raleigh, NC: State Center for Health and Environmental Statistics, North Carolina Department of Environment, Health, and Natural Resources. November.
23. North Carolina Central Registry. 1992. *North Carolina health statistics pocket guide.* Raleigh, NC: State Center for Health and Environmental Statistics, North Carolina Department of Environment, Health, and Natural Resources. December.
24. Gibbons JD, Pratt PD. 1975. P-values: Interpretation and methodology. *American Statistician* 29:20–5.